The Annals of Statistics
1990, Vol. 18, No. 1, 281–302

NONPARAMETRIC ESTIMATION OF OPTIMAL PERFORMANCE CRITERIA IN QUALITY ENGINEERING

BY R. J. CARROLL AND PETER HALL

Texas A & M University and Australian National University

Box and Leon, Shoemaker and Kackar have discussed the problem of closeness to target in quality engineering. If the mean response $f(x, z)$ depends on (x, z), the variance function is a PERMIA if it is $g(z)$, i.e., depends only on z. The goal is to find (x_0, z_0) which minimizes variance while achieving a target mean value. We pose and answer the question: For given smoothness assumptions about f and g, how accurately can we estimate x_0 and z_0? As part of the investigation, we also find optimal rates of convergence for estimating f, g and their derivatives.

1. Introduction. We investigate estimation of optimal policies in what Box (1988) calls the problem of "closeness to target" in quality engineering; see also Leon, Shoemaker and Kackar (1988) and Taguchi and Wu (1985). System variability is governed by a control factor z, so that observations have variance function $g(z)$. System mean is governed not only by the control factor z but also by a signal factor x, so that observations have mean function $f(x, z)$. In the terminology of Leon, Shoemaker and Kackar (1988), the variance function $g(z)$ is a PERMIA. As in Box (1988), the goal is to find the control setting z_0 which minimizes g and to find the signal setting x_0 for which $f(x_0, z_0) = \tau_0$, where τ_0 is a prespecified target value.

For example, consider a production line producing extruded plastic parts, whose mean length should equal τ_0. Actual mean length $f(x, z)$ is influenced by the length x into which the hot parts are cut and by the temperature z (or any quantity, such as viscosity, which is a monotone function of temperature) of the parts when they are cut. We would like to choose the pair (x, z) such that mean length equals τ_0 and variance of length is minimized. Assuming that variance of length depends only on temperature, we choose $z = z_0$ to minimize the variance $g(z)$ and then select $x = x_0$ to solve $f(x, z_0) = \tau_0$. We assume that g has a unique minimum and that $f(\cdot, z)$ has a unique minimum for each fixed z.

In practice, f and g would usually be unknown, and so we sample a variety of signal factors and control factors to produce estimators $\hat{f}$ and $\hat{g}$ of f and g, respectively. Choose $\hat{z}_0$ to minimize $\hat{g}$, and given $\hat{z}_0$, choose $\hat{x}_0$ so that $\hat{f}(\hat{x}_0, \hat{z}_0) = \tau_0$. Interest in this paper focuses on the case where f and g cannot be specified parametrically. We pose and answer the question: For given smoothness assumptions about f and g, how accurately can we estimate x_0 and z_0?

Practical interest usually does centre on estimation of x_0 and z_0, because the production line will be operated for an indefinite (or at least, unspecified) period

Received September 1987; revised November 1988.
AMS 1980 *subject classifications*. Primary 62G05; secondary 62G20.
Key words and phrases. Nonparametric regression, performance measure, PERMIA, quality control, Taguchi's method, variance function estimation.

of time with specific settings of these parameters. Nevertheless one can envisage alternative prescriptions for optimality. For example, if we were to focus on the loss associated with a given choice $(\hat{x}, \hat{z})$, we might wish to choose them to minimize

$$E\{f(\hat{x}, \hat{z}) - \tau_0\}^2 \quad \text{or} \quad E|f(\hat{x}, \hat{z}) - \tau_0| \quad \text{or} \quad P\{|f(\hat{x}, \hat{z}) - \tau_0| > c\}.$$

Some insight into the problem may be obtained by simple Taylor expansion, as follows. Assume f and g have one and two continuous derivatives, respectively. Then it is reasonable to suppose $\hat{f}$ and $\hat{g}$ to satisfy those smoothness conditions. Since $g'(z_0) = \hat{g}'(\hat{z}_0) = 0$, then

$$0 = \hat{g}'(\hat{z}_0) = \hat{g}'(z_0) + (\hat{z}_0 - z_0)\hat{g}''(\hat{z}_0^{\dagger}) = \hat{g}'(z_0) - g'(z_0) + (\hat{z}_0 - z_0)\hat{g}''(\hat{z}_0^{\dagger}),$$

where $\hat{z}_0^{\dagger}$ lies between z_0 and $\hat{z}_0$. Therefore,

$$\hat{z}_0 - z_0 = -\{\hat{g}'(z_0) - g'(z_0)\}/\hat{g}''(\hat{z}_0^{\dagger}). \tag{1.1}$$

Likewise, since $f(x_0, z_0) = \hat{f}(\hat{x}_0, \hat{z}_0) = \tau_0$, then

$$\begin{aligned}\tau_0 = \hat{f}(\hat{x}_0, \hat{z}_0) &= \hat{f}(x_0, \hat{z}_0) + (\hat{x}_0 - x_0)\hat{f}^{(1,0)}(\hat{x}_0^*, \hat{z}_0)\\ &= \tau_0 + \hat{f}(x_0, z_0) - f(x_0, z_0) + (\hat{z}_0 - z_0)\hat{f}^{(0,1)}(x_0, \hat{z}_0^*)\\ &\quad + (\hat{x}_0 - x_0)\hat{f}^{(1,0)}(\hat{x}_0^*, \hat{z}_0),\end{aligned}$$

where $\hat{x}_0^*$ lies between x_0 and $\hat{x}_0$, and $\hat{z}_0^*$ lies between z_0 and $\hat{z}_0$. Therefore,

$$\begin{aligned}\hat{x}_0 - x_0 = &-\{\hat{f}(x_0, z_0) - f(x_0, z_0)\}/\hat{f}^{(1,0)}(\hat{x}_0^*, \hat{z}_0)\\ &-(\hat{z}_0 - z_0)\hat{f}^{(0,1)}(x_0, \hat{z}_0^*)/\hat{f}^{(1,0)}(\hat{x}_0^*, \hat{z}_0).\end{aligned} \tag{1.2}$$

From equations (1.1) and (1.2) we conclude that (i) if $g^{(2)}(z_0)$ is nonzero, then z_0 can be estimated with the same accuracy as $g^{(1)}(z_0)$, and (ii) if $f^{(1,0)}(x_0, z_0)$, $f^{(0,1)}(x_0, z_0)$ and $g^{(2)}(z_0)$ are nonzero, then x_0 can be estimated with the worst of the accuracies with which $f(x_0, z_0)$ and $g^{(1)}(z_0)$ can be estimated. In the pathological event that one or other of these functions should be zero, higher-order Taylor expansions must be investigated.

Thus, estimation of x_0 and z_0 reduces to estimation of f, g and derivatives of those functions. Inference about the mean f is a classic nonparametric regression problem, but not so inference about the variance g. There, interest centres on the effect which not knowing f has on our ability to estimate g. The problem of variance function estimation in the presence of an unknown f would be one of semiparametric inference if we had a parametric model for f.

We now discuss convergence rates obtainable from (1.1) and (1.2). Suppose f has ν_1 derivatives and g has ν_2 derivatives. We allow ν_1 and ν_2 to be arbitrary positive numbers, since fractional derivatives may be expressed in terms of Lipschitz conditions. (See the second paragraph of Section 2 for definitions.) The argument leading to (1.1) and (1.2) requires at least one derivative of f and two derivatives of g, and so we assume here that $\nu_1 > 1$ and $\nu_2 > 2$. In Sections 2 and 4 we shall use (1.1) and (1.2) to show that kernel-type estimators achieve convergence rates

$$(1.3) \qquad |\hat{x}_0 - x_0| = O_p\{\max(N^{-\nu/2(\nu_1+1)}, N^{-(\nu_2-1)/(2\nu_2+1)})\},$$

$$(1.4) \qquad |\hat{z}_0 - z_0| = O_p(N^{-(\nu_2-1)/(2\nu_2+1)}),$$

where N denotes the number of pairs of signal factors and control factors in our sample. The first contribution to the right-hand side of (1.3) is due to the possible effect of not knowing f. When $\nu_1 > 1$ and $\nu_2 > 2$, not knowing f has no effect on the accuracy with which we can estimate z_0, but does influence the accuracy with which we can estimate x_0. Of course, x_0 is defined in terms of f, and so this fact occasions no surprise. A necessary and sufficient condition for the right-hand side of (1.3) to equal $O_p(N^{-(\nu_2-1)/(2\nu_2+1)})$, and so for there to be no penalty in not knowing f, is $\nu_1 \geq (2/3)(\nu_2 - 1)$.

That there should be some sort of "smoothness threshold" at which not knowing f begins to worsen performance, is not altogether surprising. If f is sufficiently smooth, then by working with high-order differences of data points, the influence of f on variance estimates can be rendered negligible. However, it seems difficult to give a simple, cogent argument describing why the threshold takes the value which it does.

We shall prove in Section 3 that the rates of convergence described by (1.3) and (1.4) are optimal, in the sense that under the stated smoothness assumptions, no nonparametric estimator can achieve faster rates.

Result (1.2), which leads to rates of convergence for estimates of z_0, requires only $\nu_2 > 2$ and $\nu_1 > 0$. We shall show that in this general circumstance, the best achievable rate of convergence of any estimator of z_0 is

$$(1.5) \qquad |\hat{z}_0 - z_0| = O_p\{\max(N^{-(\nu_2-1)/(2\nu_2+1)}, N^{-\nu_1(\nu_2-1)/\{(\nu_1+1)\nu_2\}})\}.$$

For small ν_1, this rate is inferior to that described by (1.4) unless $\nu_1(\nu_2 - 1)/\{(\nu_1 + 1)\nu_2\} \geq (\nu_2 - 1)/(2\nu_2 + 1)$; that is, unless $\nu_1 \geq \nu_2/(\nu_2 + 1)$. Of course, the latter inequality is always satisfied when $\nu_1 > 1$, and in that case (1.4) and (1.5) are identical.

In some respects the problem of estimating z_0 is a little like that of estimating the mode of a density function. See Parzen (1962), Eddy (1980) and Müller (1984) for an account of the latter problem. However, aside from the common feature that the turning point of a nonparametrically determined function is sought in both cases, there are several important dissimilarities. Not least of these is the way in which the variance estimator is defined, in terms of squared residuals which themselves involve a nonparametric curve estimator. There does not seem to be any hope of simplifying our argument by appealing to results on mode estimation.

Most of our attention will be devoted to the case of an experiment of fixed design, defined by model (2.1) in Section 2. Fixed design is more realistic than random design in most control contexts and is amenable to complete asymptotic analysis. Section 4 will outline analogous results in the random design case. Some of this work has a counterpart in heteroscedastic, nonparametric regression and will be discussed elsewhere in that context.

In some applications, our model (2.1) applies only after a data transformation of the response variable. Our discussion still applies for the closeness-to-target problem, by using approximations suggested by Box (1987) [see his equation (15)].

There has recently been work done on variance estimation in nonparametric regression, although unrelated to the problem of variance function minimization. It includes Gasser, Sroka and Jenner (1986), Buckley, Eagleson and Silverman (1988) and Hall and Carroll (1989).

2. Fixed design case. In the fixed design case our model is

$$Y_{ij} = f(i/n, j/n) + g(j/n)^{1/2}\varepsilon_{ij}, \qquad 1 \le i,\ j \le n, \tag{2.1}$$

where the ε_{ij}'s are independent with zero means, unit variances and uniformly bounded fourth moments. We observe the data set $\{Y_{ij},\ 1 \le i,\ j \le n\}$, and wish to estimate f, g and their derivatives. Note that there are $N \equiv n^2$ observations, not n; this is important when comparing our results with those in classical nonparametric regression problems.

Let $\nu > 0$, and write $\langle \nu \rangle$ for the largest integer strictly less than ν. A univariate function g is said to be ν-smooth if it has $\langle \nu \rangle$ bounded derivatives and if $g^{(\langle \nu \rangle)}$ satisfies a Lipschitz condition of order $\nu - \langle \nu \rangle$:

$$|g^{(\langle \nu \rangle)}(x) - g^{(\langle \nu \rangle)}(y)| \le C|x - y|^{\nu - \langle \nu \rangle}$$

for all $x, y \in (0,1)$. A bivariate function f is said to be ν-smooth if $f^{(i,j)}(x, y)$ exists and is bounded for all $i \ge 0$, $j \ge 0$ satisfying $i + j \le \langle \nu \rangle$ and if

$$|f^{(i,j)}(u, v) - f^{(i,j)}(x, y)| \le C\big(|u - x|^{\nu - \langle \nu \rangle} + |v - y|^{\nu - \langle \nu \rangle}\big)$$

for all $u, v, x, y \in (0,1)$ and all $i \ge 0$, $j \ge 0$ satisfying $i + j = \langle \nu \rangle$. We assume that in model (2.1), the bivariate mean function f is ν_1-smooth and the univariate variance function g is ν_2-smooth.

Our estimates of f and g are based on fixed-design analogues of kernel sequences which may be defined as follows. Given $0 < h_1, h_2 < 1$ and nonnegative integers r, s and t, let $\{a_k(h_1),\ -\infty < k < \infty\}$, $\{b_k(h_1),\ -\infty < k < \infty\}$ and $\{c_k(h_2),\ -\infty < k < \infty\}$ be sequences of constants satisfying

$$|a_k| \le Ch_1^{r+1}, \qquad |b_k| \le Ch_1^{s+1}, \qquad |c_k| \le Ch_2^{t+1},$$

$$a_k = b_k = 0 \quad \text{if } |k| > Ch_1^{-1}, \qquad c_k = 0 \quad \text{if } |k| > Ch_2^{-1},$$

$$\begin{aligned}
&\sum_k k^i a_k = \begin{cases} r! & \text{if } i = r, \\ 0 & \text{if } 0 \le i \le \langle \nu_1 \rangle \text{ and } i \ne r, \end{cases} \\
&\sum_k k^i b_k = \begin{cases} s! & \text{if } i = s, \\ 0 & \text{if } 0 \le i \le \langle \nu_1 \rangle \text{ and } i \ne s, \end{cases} \\
&\sum_k k^i c_k = \begin{cases} t! & \text{if } i = t, \\ 0 & \text{if } 0 \le i \le \langle \nu_2 \rangle \text{ and } i \ne t. \end{cases}
\end{aligned} \tag{2.2}$$

The constant C does not depend on h_1 or h_2.

To construct $\{a_k\}$, for example, let K be a compactly supported, real-valued, r-times continuously differentiable function satisfying $\int u^i K(u)\,du = 1$ if $i = 0$ and $= 0$ if $1 \le i \le \langle \nu_1 \rangle$. Put $L(u) \equiv (-1)^r K^{(r)}(u)$. Then $\int u^i L(u)\,du = r!$ if $i = r$ and $= 0$ if $0 \le i \le \langle \nu_1 \rangle$ and $i \ne r$. A slight adjustment of L, taking account of errors in series approximations to integrals and giving the function L_1, say, ensures that $a_k \equiv h_1^{r+1} L_1(h_1 k)$ has the desired properties.

Our estimator of $f^{(r,s)}$ is

$$\hat{f}^{(r,s)}(i/n, j/n) \equiv n^{r+s} \sum_k \sum_l a_k b_l Y_{i+k,\, j+l}, \tag{2.3}$$

where Y_{ij} is defined to be zero if one or other of i, j is less than 1 or greater than n. Basic properties of $\hat{f}^{(r,s)}$ are described by the following theorem.

THEOREM 2.1. *Assume f is ν_1-smooth, $\nu_1 \ge r + s$, g is bounded, $\sup E(\varepsilon_{ij}^2) < \infty$ and $h_1 = h_1(n)$ satisfies $h_1 \to 0$ and $nh_1 \to \infty$. Then, for each $0 < \delta < \frac{1}{2}$,*

$$\sup_{\delta n < i,\, j < (1-\delta)n} |E\hat{f}^{(r,s)}(i/n, j/n) - f^{(r,s)}(i/n, j/n)| = O\{(nh_1)^{-(\nu_1 - r - s)}\}, \tag{2.4}$$

$$\sup_{1 \le i,\, j \le n} \operatorname{var}\{\hat{f}^{(r,s)}(i/n, j/n)\} = O\{(nh_1)^{2(r+s)} h_1^2\}. \tag{2.5}$$

REMARK 2.1. Given any $(x, y) \in (0,1)^2$, we may define $\hat{f}^{(r,s)}(x, y)$ by linear interpolation among the four vertices of the integer square containing (x, y). It is easily shown that analogues of (2.4) and (2.5) hold for this "more general" estimator:

$$\sup_{\delta < x,\, z < 1-\delta} |E\hat{f}^{(r,s)}(x, z) - f^{(r,s)}(x, z)| = O\{(nh_1)^{-(\nu_1 - r - s)}\},$$

$$\sup_{0 < x,\, z < 1} \operatorname{var}\{\hat{f}(x, z)\} = O\{(nh_1)^{2(r+s)} h_1^2\}.$$

REMARK 2.2. It follows from Theorem 2.1 that the mean squared error of $\hat{f}^{(r,s)}$ is

$$\begin{aligned} &E\{\hat{f}^{(r,s)}(i/n, j/n) - f^{(r,s)}(i/n, j/n)\}^2 \\ &\quad = O\{(nh_1)^{-2(\nu_1 - r - s)} + (nh_1)^{2(r+s)} h_1^2\}, \end{aligned} \tag{2.6}$$

uniformly in $\delta n \le i$, $j \le (1 - \delta)n$. The order of magnitude of the right-hand side is minimized at $O(n^{-2(\nu_1 - r - s)/(\nu_1 + 1)}) = O(N^{-(\nu_1 - r - s)/(\nu_1 + 1)})$ by taking $h_1 = n^{-\nu_1/(\nu_1 + 1)}$. By modifying techniques of Stone (1980) we may show that the rate $O(N^{-(\nu_1 - r - s)/(\nu_1 + 1)})$ is optimal in a minimax sense, where the maximum is over the class of ν_1-smooth functions having a given constant C in the Lipschitz condition and in bounds on derivatives.

If we knew f, we could form the "true" residuals $r_{ij} \equiv Y_{ij} - f(i/n, j/n) = g(j/n)^{1/2} \varepsilon_{ij}$ and construct an estimator $\tilde{g}^{(t)}$ of $g^{(t)}$ as follows:

$$\tilde{g}^{(t)}(j/n) \equiv n^{t-1} \sum_{i=1}^{n} \sum_k c_k r_{i,\, j+k}^2. \tag{2.7}$$

Here r_{ij} is defined to be zero if $j < 1$ or $j > n$, and $\{c_k\}$ is as in (2.2). An argument similar to that employed to prove Theorem 2.1 may be used to establish:

THEOREM 2.2. *Assume g is ν_2-smooth, $\nu_2 > t$, $E(\varepsilon_{ij}^4)$ is uniformly bounded and $h_2 = h_2(n)$ satisfies $h_2 \to 0$ and $nh_2 \to \infty$. Then for each $0 < \delta < \frac{1}{2}$,*

$$\sup_{\delta n < j < (1-\delta)n} |E\tilde{g}^{(t)}(j/n) - g^{(t)}(j/n)| = O\{(nh_2)^{-(\nu_2 - t)}\}$$

$$\sup_{1 \le j \le n} \operatorname{var}\{\tilde{g}^{(t)}(j/n)\} = O\{(nh_2)^{2t-1}h_2^2\}.$$

REMARK 2.3. It follows from Theorem 2.2 that the mean squared error of $\tilde{g}^{(t)}$ satisfies

$$E\{\tilde{g}^{(t)}(j/n) - g^{(t)}(j/n)\}^2 = O\{(nh_2)^{-2(\nu_2 - t)} + (nh_2)^{2t-1}h_2^2\}. \tag{2.8}$$

The right-hand side here is minimized by taking $h_2 = n^{-(2\nu_2 - 1)/(2\nu_2 + 1)}$, giving a mean square error of $O(n^{-4(\nu_2 - t)/(2\nu_2+1)}) = O(N^{-2(\nu_2 - t)/(2\nu_2 + 1)})$. Again, this rate is optimal if f is known. However, we pay a penalty for not knowing f, as Theorem 2.3 shows.

Replace the true residual r_{ij} by its estimate $\hat{r}_{ij} \equiv Y_{ij} - \hat{f}(i/n, j/n)$, giving rise to the following practical estimator of $g^{(t)}$:

$$\hat{g}^{(t)}(j/n) \equiv n^{t-1} \sum_{i=1}^{n} \sum_k c_k \hat{r}_{i,j+k}^2. \tag{2.9}$$

THEOREM 2.3. *Assume f is ν_1-smooth, g is ν_2-smooth, $\nu_2 > t$, $E(\varepsilon_{ij}^4)$ is uniformly bounded and $h_i = h_i(n)$ satisfies $h_i \to 0$ and $nh_i \to \infty$ for $i = 1, 2$. Then for each $0 < \delta < \frac{1}{2}$,*

$$\begin{aligned} &\sup_{\delta n < j < (1-\delta)n} E\{\hat{g}^{(t)}(j/n) - g^{(t)}(j/n)\}^2 \\ &\quad = O\Big[\{(nh_2)^{-2(\nu_2 - t)} + (nh_2)^{2t-1}h_2^2\} + (nh_2)^{2t}\{(nh_1)^{-2\nu_1} + h_1^2\}^2\Big]. \end{aligned} \tag{2.10}$$

REMARK 2.4. The order of the mean squared error of $\hat{g}^{(t)}$ is that of the mean squared error of $\tilde{g}^{(t)}$, plus $(nh_2)^{2t}$ times the square of the mean squared error of $\hat{f}$; compare (2.8) and (2.10), noting result (2.6) for $r = s = 0$. The additional term represents the penalty in not knowing f when estimating $g^{(t)}$.

REMARK 2.5. The value of h_1 which minimizes the order of the second term on the right-hand side of (2.10) is $h_1 = h_1^* \equiv n^{-\nu_1/(\nu_1+1)}$. Using this value of h_1, we find that

$$\begin{aligned} &E\{\hat{g}^{(t)}(j/n) - g^{(t)}(j/n)\}^2 \\ &\quad = O\Big[\{(nh_2)^{-2(\nu_2 - t)} + (nh_2)^{2t-1}h_2^2\} + (nh_2)^{2t}n^{-4\nu_1/(\nu_1+1)}\Big]. \end{aligned} \tag{2.11}$$

The value of h_2 which minimizes the order of $A(h_2) \equiv (nh_2)^{-2(\nu_2 - t)} + (nh_2)^{2t-1}h_2^2$ is $h_2 = h_2^* \equiv n^{-(2\nu_2 - 1)/(2\nu_2 + 1)}$ and $A(h_2^*) = 2n^{-4(\nu_2 - t)/(2\nu_2 + 1)}$. Furthermore, $(nh_2^*)^{2t}n^{-4\nu_1/(\nu_1+1)} \le A(h_2^*)$ if and only if

$$\nu_1 \ge \nu_2/(\nu_2 + 1). \tag{2.12}$$

Therefore, when (2.12) is true, the term involving h_1 on the right-hand side of (2.10) does not influence the convergence rate of the optimally constructed version of $\hat{g}^{(t)}$, and for $h_1 = h_1^*$ and $h_2 = h_2^*$,

$$E\{\hat{g}^{(t)}(j/n) - g^{(t)}(j/n)\}^2 = O(n^{-4(\nu_2 - t)/(2\nu_2+1)}).$$

This is the same as the best rate of convergence of $\tilde{g}^{(t)}$; see Remark 2.3.

REMARK 2.6. If (2.12) fails, then there is a cost to estimating f. An optimal balance among terms on the right-hand side of (2.11) is achieved by making $(nh_2)^{-2(\nu_2 - t)}$ the same size as $(nh_2)^{2t}n^{-4\nu_1/(\nu_1+1)}$. This is, take $h_2 = h_2^{**} \equiv n^{\{2\nu_1 - \nu_2(\nu_1+1)\}/\{(\nu_1+1)\nu_2\}}$, in which case

$$E\{\hat{g}^{(t)}(j/n) - g^{(t)}(j/n)\}^2 = O(n^{-4\nu_1(\nu_2 - t)/\{(\nu_1+1)\nu_2\}}).$$

REMARK 2.7. Note that h_2^{**} [the optimal version of h_2 when (2.12) fails] is different from h_2^* [the optimal h_2 when (2.12) holds]. Also, none of h_1^*, h_2^* and h_2^{**} depends on t.

REMARK 2.8. We may summarize the main points made during Remarks 2.5 and 2.6 by stating that if $\hat{g}^{(t)}$ is constructed using $h_1 = h_1^*$ and $h_2 = h_2^*$ [if (2.12) holds] or $h_2 = h_2^{**}$ [if (2.12) fails], then

$$\begin{aligned} &E\{\hat{g}^{(t)}(j/n) - g^{(t)}(j/n)\}^2 \\ &\quad = O\{\max(n^{-4(\nu_2 - t)/(2\nu_2+1)}, n^{-4\nu_1(\nu_2 - t)/\{(\nu_1+1)\nu_2\}})\}. \end{aligned} \tag{2.13}$$

The term involving only ν_2 dominates the right-hand side here if (2.12) holds, while the other term dominates if (2.12) fails. We shall show in Section 3 that the rate of convergence described by (2.13) is optimal in a minimax sense.

To solve the first part of our control problem, we need to estimate that value z_0 which minimizes g. If g has a continuous derivative, then this amounts to estimating the solution z_0 of the equation $g^{(1)}(z) = 0$. A potential estimator $\hat{g}^{(1)}(z)$ of $g^{(1)}(z)$ may be obtained by interpolating among values of $\hat{g}^{(1)}(j/n)$, defined at (2.9). However, this approach results in a very rough estimator, without even a single continuous derivative. There are several ways of deriving a smoother estimator. One is to derive $\hat{g}^{(2)}(z)$ by linearly interpolating among values of $\hat{g}^{(2)}(j/n)$, and then estimate $g^{(1)}$ by integrating $\hat{g}^{(2)}$. This we do below.

Define $\hat{g}^{(1)}(j/n)$ and $\hat{g}^{(2)}(j/n)$ as at (2.9), construct $\hat{g}^{(2)}(z)$ by linearly interpolating among points $\hat{g}^{(2)}(j/n)$ and, for an arbitrary j_0 satisfying $j_0 \sim n\alpha$

for some $0 < \alpha < 1$, put

$$\bar{g}^{(1)}(z) \equiv \hat{g}^{(1)}(j_0/n) + \int_{j_0/n}^{z} \hat{g}^{(2)}(u)\,du, \qquad 0 < z < 1.$$

This will be our estimator of $g^{(1)}(z)$. It is continuously differentiable, with derivative $(\bar{g}^{(1)})'(z) = \hat{g}^{(2)}(z)$, and is a quadratic interpolation of an estimator "like" $\hat{g}^{(1)}$. It shares the mean squared error properties of $\hat{g}^{(1)}$, as follows.

THEOREM 2.4. *Assume the conditions of Theorem* 2.3, *with* $t = 1$. *Then for each* $0 < \delta < \frac{1}{2}$,

$$\sup_{\delta < z < 1-\delta} E\{\bar{g}^{(1)}(z) - g^{(1)}(z)\}^2 = O\Big[\big\{(nh_2)^{-2(\nu_2-1)} + nh_2^3\big\} + (nh_2)^2\big\{(nh_1)^{-2\nu_1} + h_1^2\big\}^2\Big]. \tag{2.14}$$

REMARK 2.9. Note that the right-hand sides of (2.10) (for $t = 1$) and (2.14) are identical.

REMARK 2.10. The conditions in Theorem 2.4 do not require the existence of a second derivative of g, even though $\hat{g}^{(2)}$ is used in the construction of $\bar{g}^{(1)}$. We need only assume $\nu_2 > 1$; of course, $\hat{g}^{(2)}$ is well-defined without any smoothness assumptions, being given by formula (2.9).

We are now in a position to solve the first part of our control problem. Let $\hat{z}_0$ be any solution of the equation $\bar{g}^{(1)}(\hat{z}_0) = 0$, and z_0 be the unique solution of $g^{(1)}(z_0) = 0$. Then

$$0 = \bar{g}^{(1)}(\hat{z}_0) = \bar{g}(z_0) + (\hat{z}_0 - z_0)\hat{g}^{(2)}\{z_0 + \hat{\theta}(\hat{z}_0 - z_0)\}, \tag{2.15}$$

where $0 \le \hat{\theta} \le 1$. Assume g is ν_2-smooth for some $\nu_2 > 2$. Then $g^{(2)}$ is well-defined and continuous. Suppose that for an integer $l \ge 1$, $4l$th moments of the errors ε_{ij} are uniformly bounded. Then the argument leading to Theorem 2.3 may be generalized to prove that for each $0 < \delta < \frac{1}{2}$,

$$\sup_{\delta n < j < (1-\delta)n} E\{\hat{g}^{(2)}(j/n) - g^{(2)}(j/n)\}^{2l} = O\{B_2(h_1, h_2)^l\},$$

where $B_t(h_1, h_2) \equiv \{(nh_2)^{-2(\nu_2-t)} + (nh_2)^{2t-1}h_2^2\} + (nh_2)^{2t}\{(nh_1)^{-2\nu_1} + h_1^2\}^2$. Choose h_1, h_2 to minimize the order of $B_1(h_1, h_2)$, as described in Remark 2.8. Then $B_2(h_1, h_2) = O(n^{-b})$, where $b \equiv \min[4(\nu_2 - 2)/(2\nu_2 + 1), 4\nu_1(\nu_2 - 2)/\{(\nu_1 + 1)\nu_2\}] > 0$. If $l > 1/b$, then for each $\eta > 0$ and each $0 < \delta < \frac{1}{2}$, we have by Markov's inequality,

$$P\Big\{\sup_{\delta n < j < (1-\delta)n} |\hat{g}^{(2)}(j/n) - g^{(2)}(j/n)| > \eta\Big\} = O(n^{1-bl}) = o(1),$$

so that

$$\sup_{\delta < z < 1-\delta} |\hat{g}^{(2)}(z) - g^{(2)}(z)| = o_p(1). \tag{2.16}$$

Therefore, by (2.15), assuming that $g^{(2)}(z_0) \neq 0$,

$$\hat{z}_0 - z_0 = -\{1 + o_p(1)\}\{\bar{g}^{(1)}(z_0) - g^{(1)}(z_0)\}/g^{(2)}(z_0).$$

We conclude that $\hat{z}_0$ converges to z_0 at the same rate as $\bar{g}^{(1)}(z_0)$ converges to $g^{(1)}(z_0)$; that is,

$$|\hat{z}_0 - z_0| = O_p\{\max(n^{-2(\nu_2-1)/(2\nu_2+1)}, n^{-2\nu_1(\nu_2-1)/\{(\nu_1+1)\nu_2\}})\}. \tag{2.17}$$

This is result (1.5), announced in Section 1, and implies (1.4) when $\nu_1 > 1$.

The second part of our control problem consists of estimating the value x_0 which satisfies $f(x_0, z_0) = \tau_0$. An estimator of f is $\hat{f} = \hat{f}^{(0,0)}$, defined in (2.3) with $r = s = 0$. However, as in the case of our estimator of $g^{(1)}$, this suffers from being "too rough." Therefore we compute $\hat{f}^{(0,1)}$, $\hat{f}^{(1,0)}$ and $\hat{f}^{(1,1)}$ by linearly interpolating among values defined in (2.3), and then derive an estimator $\bar{f}$ of f by integration, as follows. Let i_0, j_0 satisfy $i_0 \sim n\alpha$, $j_0 \sim n\beta$, where $0 < \alpha, \beta < 1$, and put

$$\begin{aligned}\bar{f}(x,z) \equiv \hat{f}(i_0/n, j_0/n) &+ \int_{i_0/n}^{x} \hat{f}^{(1,0)}(u, j_0/n)\,du + \int_{j_0/n}^{z} \hat{f}^{(0,1)}(i_0/n, v)\,dv \\ &+ \int_{i_0/n}^{x} du \int_{j_0/n}^{z} \hat{f}^{(1,1)}(u,v)\,dv, \qquad 0<x,\ z<1.\end{aligned} \tag{2.18}$$

This will be our estimator of $f(x, z)$. It is continuously differentiable in both variables, satisfying

$$(\partial/\partial x)\bar{f}(x,z) = \hat{f}^{(1,0)}(x, j_0/n) + \int_{j_0/n}^{z} \hat{f}^{(1,1)}(x,v)\,dv$$

and an analogous expression for $(\partial/\partial z)\bar{f}(x, z)$. It shares the mean squared error properties of $\hat{f}^{(0,0)}$, as follows.

THEOREM 2.5. *Assume the condition of Theorem* 2.1, *with* $r = s = 0$. *Then for each* $0 < \delta < \frac{1}{2}$,

$$\sup_{\delta < z < 1-\delta} E\{\bar{f}(x,z) - f(x,z)\}^2 = O\{(nh_1)^{-2\nu_1} + h_1^2\}. \tag{2.19}$$

REMARK 2.11. Note that the right-hand sides of (2.14) (for $r = s = 0$) and (2.19) are identical.

REMARK 2.12. Theorem 2.5 does not require the existence of any derivative of f, even though numerical values of $\hat{f}^{(0,1)}$, $\hat{f}^{(1,0)}$ and $\hat{f}^{(1,1)}$ are used in the construction of $\bar{f}$.

We are now in a position to solve the second part of our control problem. Suppose f is ν_1-smooth, where $\nu_1 > 1$. Then $f^{(0,1)}$ and $f^{(1,0)}$ are well-defined and continuous. Assume $f^{(0,1)}(x_0, z_0) \neq 0 \neq f^{(1,0)}(x_0, z_0)$. Define $\bar{f}$ as at (2.18) and write $\bar{f}^{(i,j)}(x, z)$ for $(\partial/\partial x)^i(\partial/\partial z)^j \bar{f}(x, z)$. Choose $h_1 = n^{-\nu_1/(\nu_1+1)}$ to minimize the order of $(nh_1)^{-2\nu_1} + h_1^2$. Then by (2.19),

$$|\bar{f}(x_0, z_0) - f(x_0, z_0)| = O_p(n^{-\nu_1/(\nu_1+1)}). \tag{2.20}$$

Suppose that for an integer $l \geq 1$, $2l$th moments of the errors ε_{ij} are uniformly bounded. The argument leading to (2.16) may be modified to show that if l is sufficiently large, then for each $0 < \delta < \frac{1}{2}$,

$$\sup_{\delta < x,\, z < 1-\delta} |\bar{f}^{(i,j)}(x, z) - f^{(i,j)}(x, z)| = o_p(1) \tag{2.21}$$

for $(i, j) = (0,1)$ or $(1,0)$. Using the Taylor expansion which produced (1.2) we may now deduce that

$$\begin{aligned}\hat{x}_0 - x_0 = &-\{1 + o_p(1)\}\{\bar{f}(x_0, z_0) - f(x_0, z_0)\}/f^{(1,0)}(x_0, z_0)\\ &-\{1 + o_p(1)\}(\hat{z}_0 - z_0)f^{(0,1)}(x_0, z_0)/f^{(1,0)}(x_0, z_0).\end{aligned}$$

We conclude that the rate of convergence of $\hat{x}_0$ to x_0 is the worst of the rates of convergence of $\bar{f}(x_0, z_0)$ to $f(x_0, z_0)$ and of $\hat{z}_0$ to z_0. By (2.17) and (2.20), this is

$$\begin{aligned}|\hat{x}_0 - x_0| &= O_p\{\max(n^{-\nu_1/(\nu_1+1)}, n^{-2(\nu_2-1)/(2\nu_2+1)}, n^{-2\nu_1(\nu_2-1)/\{(\nu_1+1)\nu_2\}})\}\\ &= O_p\{\max(n^{-\nu_1/(\nu_1+1)}, n^{-2(\nu_2-1)/(2\nu_2+1)})\},\end{aligned}$$

the second identity following from the fact that $\nu_1 > 1$. This is result (1.3), announced in Section 1.

PROOF OF THEOREM 2.1. We begin with a lemma.

LEMMA. *Let $m \geq 0$. Suppose the bivariate function f has continuous derivatives $f^{(i,j)}$, for $i \geq 0$, $j \geq 0$ and $i + j \leq m$, on the square $[0,1]^2$. There exist numbers θ_{i1}, θ_{i2} satisfying $0 \leq \theta_{ij} \leq 1$, such that*

$$\begin{aligned}f(u_1 + \delta_1, u_2 + \delta_2) = &\sum\sum_{0 \leq i+j \leq m-1} (\delta_1^i \delta_2^j / i!j!) f^{(i,j)}(u_1, u_2)\\ &+ \sum\sum_{i+j=m} (\delta_1^i \delta_2^j / i!j!) f^{(i,j)}(u_1 + \theta_{i1}\delta_1, u_2 + \theta_{i2}\delta_2)\end{aligned}$$

whenever $u_1, u_2, u_1 + \delta_1, u_2 + \delta_2 \in [0,1]$.

To prove the lemma, write $f(u_1 + \delta_1, u_2 + \delta_2) = \{f(u_1 + \delta_1, u_2 + \delta_2) - f(u_1, u_2 + \delta_2)\} + f(u_1, u_2 + \delta_2)$, and repeatedly apply the univariate version of Taylor's theorem with remainder.

To prove (2.4), put $m \equiv \langle \nu_1 \rangle$ and apply the lemma, obtaining for integer α and β,

$$E\left\{ \hat{f}^{(r,s)}\left(\frac{\alpha}{n}, \frac{\beta}{n}\right) - f^{(r,s)}\left(\frac{\alpha}{n}, \frac{\beta}{n}\right)\right\}$$

$$= n^{r+s} \sum_k \sum_l a_k b_l \sum\sum_{i+j=m} \frac{(k/n)^i (l/n)^j}{i!j!}$$

$$\times \left\{ f^{(i,j)}\left(\frac{\alpha + \theta_{i1}k}{n}, \frac{\beta + \theta_{i2}l}{n}\right) - f^{(i,j)}\left(\frac{\alpha}{n}, \frac{\beta}{n}\right)\right\}$$

$$= O\left\{ n^{r+s} \sum_k \sum_l \sum\sum_{i+j=m} \left|\left(\frac{k}{n}\right)^i \left(\frac{l}{n}\right)^j a_k b_l\right| \left(\left|\frac{k}{n}\right|^{\nu_1 - m} + \left|\frac{l}{n}\right|^{\nu_1 - m}\right)\right\}$$

$$= O\{(nh_1)^{r+s-\nu_1}\}.$$

To prove (2.5), observe that

$$\operatorname{var}\left\{ \hat{f}^{(r,s)}\left(\frac{i}{n}, \frac{j}{n}\right)\right\} = O\left\{n^{2(r+s)}\left(\sum a_k^2\right)\left(\sum b_k^2\right)\right\} = O\left\{(nh_1)^{2(r+s)} h_1^2\right\}. \quad \square$$

PROOF OF THEOREM 2.3. Take $r = s = 0$, in which case we may assume $a_l = b_l$, and our estimator of f is

$$\hat{f}(i/n, j/n) = \sum_{l_1} \sum_{l_2} a_{l_1} a_{l_2} Y_{i+l_1, j+l_2}.$$

Put

$$\Delta_{ij} \equiv \sum_{l_1} \sum_{l_2} a_{l_1} a_{l_2} g\{(j + l_2)/n\}^{1/2} \varepsilon_{i+l_1, j+l_2}$$

and

$$B_{ij} \equiv \sum_{l_1} \sum_{l_2} a_{l_1} a_{l_2} f\{(i + l_1)/n, (j + l_2)/n\} - f(i/n, j/n).$$

In this notation, $\hat{r}_{ij} = r_{ij} - \Delta_{ij} - B_{ij}$, so that $n^{1-t}\hat{g}^{(t)}(j/n) = n^{1-t}\tilde{g}^{(t)}(j/n) - 2A_j + B_j$, where

$$A_j \equiv \sum_{i=1}^{n} \sum_k (\Delta_{i,j+k} + B_{i,j+k}) r_{i,j+k} c_k, \qquad B_j \equiv \sum_{i=1}^{n} \sum_k (\Delta_{i,j+k} + B_{i,j+k})^2 c_k.$$

Therefore, in view of (2.8), it suffices to prove that

$$(2.22) \qquad \begin{aligned} E(A_j^2) &= O\left[(nh_2^t)^2\{(nh_1)^{-2\nu_1} + h_1^2\}^2 + h_2^{2(t+1)}\right], \\ E(B_j^2) &= O\left[(nh_2^t)^2\{(nh_1)^{-2\nu_1} + h_1^2\}^2\right]. \end{aligned}$$

Since $A_j \equiv \Sigma_i \Sigma_k (\Delta_{i, j+k} + B_{i, j+k}) g\{(j+k)/n\}^{1/2} \varepsilon_{i, j+k} c_k$, then

$$\tag{2.23} \begin{aligned} \tfrac{1}{2} E\left(A_j^2\right) \le E\left[\sum_i \sum_k \Delta_{i, j+k} \varepsilon_{i, j+k} g\{(j+k)/n\}^{1/2} c_k\right]^2 \\ + E\left[\sum_i \sum_k \varepsilon_{i, j+k} B_{i, j+k} g\{(j+k)/n\}^{1/2} c_k\right]^2. \end{aligned}$$

Now,

$$\begin{aligned} &E\left(\Delta_{i_1, j+k_1} \varepsilon_{i_1, j+k_1} \Delta_{i_2, j+k_2} \varepsilon_{i_2, j+k_2}\right) \\ &\quad = \sum_{l_1, \dots, l_4} \cdots \sum a_{l_1, \dots, l_4} \left[\prod_{\alpha=1}^{4} g\{(j + l_\alpha)/n\}\right]^{1/2} \\ &\qquad \times E\left(\varepsilon_{i_1 + l_1, j+k_1+l_2} \varepsilon_{i_1, j+k_1} \varepsilon_{i_2 + l_3, j+k_2+l_4} \varepsilon_{i_2, j+k_2}\right) \\ &\quad = O\left\{h_1^4 + h_1^2 I(i_1 = i_2, k_1 = k_2)\right\}. \end{aligned}$$

Hence, the first term on the right-hand side of (2.23) equals

$$\begin{aligned} &O\left[\sum_{i_1, i_2, k_1, k_2} \cdots \sum \left\{h_1^4 + h_1^2 I(i_1 = i_2, k_1 = k_2)\right\} h_2^{2(t+1)} I\left(|k_1|, |k_2| \le C h_2^{-1}\right)\right] \\ &\quad = O\left\{\left(n h_2^t\right)^2 h_1^4 + n h_1^2 h_2^{2t+1}\right\} = O\left\{\left(n h_2^t\right)^2 h_1^4 + h_2^{2(t+1)}\right\}. \end{aligned}$$

Since $|B_{ij}| = O\{(nh_1)^{-\nu_1}\}$, then the second term on the right-hand side of (2.23) equals

$$\begin{aligned} E\left(\varepsilon_{11}^2\right) \sum_i \sum_k B_{i, j+k}^2 g\{(j+k)/n\} c_k^2 &= O\left\{(nh_1)^{-2\nu_1} n h_2^{2t+1}\right\} \\ &= O\left\{\left(n h_2^t\right)^2 (nh_1)^{-4\nu_1} + h_2^{2(t+1)}\right\}. \end{aligned}$$

Combining estimates from (2.23) down we get the first part of (2.22). The second part follows from the fact that $|c_k| \le C h_2^{t+1} I(|k| \le C h_2^{-1})$, $E(\Delta_{ij}^4) = O(h_1^4)$ and $|B_{ij}| = O\{(nh_1)^{-\nu_1}\}$. □

PROOF OF THEOREM 2.4. If $j/n \le u \le (j+1)/n$, then

$$\hat{g}^{(2)}(n) = (nu - j)\hat{g}^{(2)}\{(j+1)/n\} + (j + 1 - nu)\hat{g}^{(2)}(j/n),$$

whence

$$\int_{j/n}^{(j+1)/n} \hat{g}^{(2)}(u)\, du = (2n)^{-1}\left[\hat{g}^{(2)}(j/n) + \hat{g}^{(2)}\{(j+1)/n\}\right].$$

Therefore, if $j/n \le z \le (j+1)/n$ and $j \ge j_0 + 2$,

$$\begin{aligned} \bar{g}^{(1)}(z) &= \hat{g}^{(1)}(j_0/n) + n^{-1} \sum_{i=j_0+1}^{j-1} \hat{g}^{(2)}(i/n) + T_1 + T_2 \\ &= \hat{g}^{(1)}(j_0/n) + \hat{g}^{*(1)}\{(j-1)/n\} - \hat{g}^{*(1)}(j_0/n) + T_1 + T_2, \end{aligned}$$

where

$$\hat{g}^{*(1)}(j/n) = \sum_{i=1}^{n} \sum_{k} d_k \hat{r}_{i,\,j+k}^2, \qquad d_k \equiv \sum_{l=0}^{\infty} c_{k+1},$$

$$T_1 \equiv \int_{j/n}^{z} \hat{g}^{(2)}(u)\,du, \qquad T_2 \equiv (2n)^{-1}\{\hat{g}^{(2)}(j_0/n) + \hat{g}^{(2)}(j/n)\}.$$

If $\{c_k\}$ satisfies condition (2.2) with $t = 2$, then $\{d_k\}$ satisfies the same condition (stated there for $\{c_k\}$) with $t = 1$. Therefore, Theorem 2.4 will follow from Theorem 2.3 if we prove that, for $i = 1$ and 2,

$$(2.24)\quad E(T_i^2) = O\Big[(nh_2)^{-2(\nu_2-1)} + nh_2^3 + (nh_2)^2\big\{(nh_1)^{-2\nu_1} + h_1^2\big\}^2\Big].$$

(The case of j values with $j \le j_0 + 1$ may be treated similarly. Note that we may not, and do not, assume existence of $g^{(2)}$.)

Observe that

$$E(T_1^2) \le n^{-2} \sup_{j/n \le u \le (j+1)/n} E\big\{\hat{g}^{(2)}(u)^2\big\} \le 2n^{-2} \max_{l=j,\,j+1} E\big\{\hat{g}^{(2)}(l/n)^2\big\}.$$

Let A_j, B_j be as in the Proof of Theorem 2.3, this time with $t = 2$. Then $\hat{g}^{(2)}(l/n) = \tilde{g}^{(2)}(l/n) + n(B_l - 2A_l)$ and (as shown during our proof of Theorem 2.3) $E(A_l^2) + E(B_l^2) = O[(nh_2^3)^2\{(nh_1)^{-2\nu_1} + h_1^2\}^2 + h_2^6]$. Also,

$$n^{-2}E\big\{\hat{g}^{(2)}(l/n)^2\big\} \le 16\big[n^{-2}E\big\{\tilde{g}^{(2)}(l/n)^2\big\} + E(A_l^2) + E(B_l^2)\big],$$

and, since $\tilde{g}^{(2)}(l/n) = n\sum_i\sum_k c_k g\{(l+k)/n\}\varepsilon_{i,\,l+k}^2$, then

$$\begin{aligned} E\big\{\tilde{g}^{(2)}(l/n)^2\big\} &= \operatorname{var}\big\{\tilde{g}^{(2)}(l/n)\big\} + \big\{E\tilde{g}^{(2)}(l/n)\big\}^2 \\ &= O\Bigg[n^3\sum_k c_k^2 + \Big|n^2\sum_k c_k g\{(l+k)/n\}\Big|^2\Bigg] \\ &= O\big\{n^3h_2^5 + (nh_2)^{2(2-\nu_2)}\big\}. \end{aligned}$$

Combining all these estimates we conclude that, for $i = 1$ and 2,

$$E(T_i^2) = O\Big[nh_2^5 + (nh_2)^{-2(\nu_2-1)}h_2^2 + (nh_2^3)^2\big\{(nh_1)^{-2\nu_1} + h_1^2\big\}^2\Big],$$

from which follows (2.24). □

3. Optimal rates of convergence. In this section we show that the convergence rates derived in Section 2 for kernel-type estimators cannot be improved upon by other estimators. Our optimality results will be in the form of "worst possible" rates computed over function classes. It is a trivial matter to obtain the same rates for our kernel-type estimators by extending arguments in Section 2. In the next paragraph we define the function classes and state the extended results.

Given positive numbers ν_1, ν_2 and B, let $\mathscr{C}_1 = \mathscr{C}_1(\nu_1, B)$ be the class of bivariate functions f on $[0,1]^2$ for which $\sup|f^{(i,j)}| \le B$ whenever $i \ge 0$, $j \ge 0$

and $i + j \le \langle \nu_1 \rangle$, and

$$|f^{(i,j)}(u, v) - f^{(i,j)}(x, y)| \le B\big(|u - x|^{\nu_1 - \langle \nu_1 \rangle} + |v - y|^{\nu_1 - \langle \nu_1 \rangle}\big)$$

whenever $u, v, x, y \in [0,1]$, $i \ge 0$, $j \ge 0$ and $i + j = \langle \nu_1 \rangle$. Let $\mathscr{C}_2 = \mathscr{C}_2(\nu_2, B)$ be the class of nonnegative univariate functions g on $[0,1]$ for which $\sup|g^{(i)}| \le B$ whenever $0 \le i \le \langle \nu_2 \rangle$ and

$$|g^{(\langle \nu_2 \rangle)}(x) - g^{(\langle \nu_2 \rangle)}(y)| \le B|x - y|^{\nu_2 - \langle \nu_2 \rangle}$$

whenever $x, y \in [0,1]$. Let $\mathscr{C}_3 = \mathscr{C}_3(B)$ be the class of nonnegative univariate functions g on $[0,1]$ such that $\sup g \le B$. Take $\hat{f}^{(r,s)}$, $\tilde{g}^{(t)}$ and $\hat{g}^{(t)}$ to be the estimators defined at (2.3), (2.7) and (2.9), respectively, calculated by linear interpolation at points which are not integer multiples of n^{-1}. (See Remark 2.1.) Assume that $\nu_1 > r + s$ and $\nu_2 > t$. For appropriate choices of the smoothing parameters h_1 and h_2, and for each $0 < \delta < \frac{1}{2}$, there exist positive constants C_1, C_2 and C_3 depending on ν_1, ν_2 and B such that

$$\sup_{f \in \mathscr{C}_1,\, g \in \mathscr{C}_3} \sup_{\delta < x,\, z < 1 - \delta} E_{f,g}\{\hat{f}^{(r,s)}(x, z) - f^{(r,s)}(x, z)\}^2 \le C_1 n^{-2(\nu_1 - r - s)/(\nu_1 + 1)},$$

$$\sup_{g \in \mathscr{C}_2} \sup_{\delta < z < 1 - \delta} E_g\{\tilde{g}^{(t)}(z) - g^{(t)}(z)\}^2 \le C_2 n^{-4(\nu_2 - t)/(2\nu_2 + 1)},$$

$$\sup_{f \in \mathscr{C}_1,\, g \in \mathscr{C}_2} \sup_{\delta < z < 1 - \delta} E_{f,g}\{\hat{g}^{(t)}(z) - g^{(t)}(z)\}^2 \le C_3 \max\big(n^{-4(\nu_2 - t)/(2\nu_2 + 1)}, n^{-4\nu_1(\nu_2 - t)/\{(\nu_1 + 1)\nu_2\}}\big).$$

These results, but without the suprema over f and g, were obtained in Remarks 2.2, 2.3 and 2.8 respectively. The methods of proof, smoothing parameters and convergence rates are exactly the same in the present uniform context.

In this section we show that, for *any* nonparametric estimators $\hat{f}^{(r,s)}$, $\tilde{g}^{(t)}$ and $\hat{g}^{(t)}$ (not just for our kernel estimators), the above inequalities may be reversed. Let $\hat{f}^{(r,s)}$ and $\hat{g}^{(t)}$ be nonparametric estimators of $f^{(r,s)}$ and $g^{(t)}$, respectively, based on model (2.1), and let $\tilde{g}^{(t)}$ be a nonparametric estimator of $g^{(t)}$, based on the true residuals $r_{ij} \equiv g(j/n)^{1/2}\varepsilon_{ij}$, $1 \le i, j \le n$. Assume that the errors ε_{ij} are independent and identically distributed as normal $N(0,1)$ and that $\nu_1 > r + s$ and $\nu_2 > t$. We claim that for any fixed $(x_0, z_0) \in (0,1)^2$ and arbitrary nonparametric estimators $\hat{f}^{(r,s)}$, $\tilde{g}^{(t)}$ and $\hat{g}^{(t)}$, there exist positive constants D_1, D_2 and D_3 such that, for large n,

$$\sup_{f \in \mathscr{C}_1,\, g \in \mathscr{C}_3} E_{f,g}\{\hat{f}^{(r,s)}(x_0, z_0) - f^{(r,s)}(x_0, z_0)\}^2 \ge D_1 n^{-2(\nu_1 - r - s)/(\nu_1 + 1)}, \tag{3.1}$$

$$\sup_{g \in \mathscr{C}_2} E_g\{\tilde{g}^{(t)}(z_0) - g^{(t)}(z_0)\}^2 \ge D_2 n^{-4(\nu_2 - t)/(2\nu_2 + 1)}, \tag{3.2}$$

$$\sup_{f \in \mathscr{C}_1,\, g \in \mathscr{C}_2} E_{f,g}\{\hat{g}^{(t)}(z_0) - g^{(t)}(z_0)\}^2 \ge D_3 \max\big(n^{-4(\nu_2 - t)/(2\nu_2 + 1)}, n^{-4\nu_1(\nu_2 - t)/\{(\nu_1 + 1)\nu_2\}}\big). \tag{3.3}$$

Results (3.1) and (3.2) may be viewed as lower bounds to convergence rates for estimation of mean functions in nonparametric regression with uniformly bounded variances. In the case of (3.2), the regression is replicated n times at each design point. Both results may be derived by modifying arguments of Stone (1980), who treats lower bounds in nonreplicated regression. Result (3.3) is more difficult to obtain and is proved in detail later in this section.

Next we turn attention to estimation of z_0, the unique element of $[0,1]$ such that $\inf g = g(z_0)$. The rate of convergence for our kernel-based estimator was described by (2.17). To extend this to a rate uniform over a function class, we must define a new function class, as follows. Fix $\nu_2 > 2$, $0 < \delta < \frac{1}{2}$ and $0 < c < \frac{1}{2}B$. Write $\mathscr{D}_2 = \mathscr{D}_2(\nu_2, \delta, B, c)$ for the class of nonnegative functions g which are in $\mathscr{C}_2(\nu_2, B)$ and which satisfy $\frac{1}{2}c \le g^{(2)}(z) \le 2c$ for $z \in [0,1]$, $g^{(1)}(z_0) = 0$ for some $z_0 \in [\delta, 1-\delta]$. It follows that each $g \in \mathscr{D}_2$ is strictly convex, with minimum attained at its unique turning point z_0. Fix $\nu_1 > 0$ and let $\mathscr{C}_1 = \mathscr{C}_1(\nu_1, B)$ be the function class defined earlier. Then if $\hat{z}_0$ is our kernel-based estimator of z_0, and if $\{a_n\}$ is a positive sequence with $a_n \to \infty$,

$$\sup_{f \in \mathscr{C}_1,\, g \in \mathscr{D}_2} P_{f,g}\Big\{|\hat{z}_0 - z_0| > a_n \max\big(n^{-2(\nu_1-1)/(2\nu_2+1)}, n^{-2\nu_1(\nu_2-1)/\{(\nu_1+1)\nu_2\}}\big)\Big\} \to 0 \tag{3.4}$$

as $n \to \infty$. (Here $\nu_1 > 0$ and $\nu_2 > 2$.) This is a version of (2.17) uniformly over function classes and is proved in the same manner as (2.17). To state a converse result, let $\hat{z}_0$ be any nonparametric estimator of z_0 and $\{a_n\}$ be any positive sequence. We claim that if (3.4) holds, then $a_n \to \infty$. An outline of the proof of this fact will be given later in this section.

Similar results for estimation of x_0 require a new class $\mathscr{D}_1$ of mean functions f. Fix $d \in (0, \frac{1}{2}B)$, $\nu_1 > 1$ and τ_0, and let $\mathscr{D}_1 = \mathscr{D}_1(\nu_1, \delta, \tau_0, B, d)$ be the class of functions f which are in $\mathscr{C}_2(\nu_2, B)$, which satisfy $\frac{1}{2}d \le |f^{(0,1)}(x,z)|, |f^{(1,0)}(x,z)| \le 2d$ for $(x,z) \in [0,1]^2$, and which are such that for each $z \in [\delta, 1-\delta]$ the equation $f(x,z) = \tau_0$ has a unique solution $x(z)$. Then if $\hat{x}_0$ is our kernel-based estimator of $x_0 = x(z_0)$, and if $\{a_n\}$ is a positive sequence with $a_n \to \infty$,

$$\sup_{f \in \mathscr{D}_1,\, g \in \mathscr{D}_2} P_{f,g}\Big\{|\hat{x}_0 - x_0| > a_n \max\big(n^{-\nu_1/(\nu_1+1)}, n^{-2(\nu_2-1)/(2\nu_2+1)}\big)\Big\} \to 0 \tag{3.5}$$

as $n \to \infty$. (Here $\nu_1 > 1$ and $\nu_2 > 2$.) Conversely, if $\hat{x}_0$ is any nonparametric estimator of x_0, if $\{a_n\}$ is a positive sequence and if (3.5) holds, then $a_n \to \infty$.

We conclude this section with a detailed proof of (3.3), and sketches of proofs of the rates of convergence described by (3.4) and (3.5).

PROOF OF (3.3). It is notationally simpler to assume a regular design on the square $[-1,1]^2$ instead of on $[0,1]^2$ and to take $x_0 = 0$. There is no loss of generality in confining attention to this situation, and so we suppose instead of model (2.1) that $Y_{ij} = f(i/n, j/n) + g(j/n)^{1/2}\varepsilon_{ij}$, $-n \le i,\ j \le n$, where the ε_{ij}'s are i.i.d. $N(0,1)$. Define the function classes $\mathscr{C}_1$ and $\mathscr{C}_2$ on $[-1,1]$ instead of on $[0,1]$.

In the case $\nu_1 \ge \nu_2/(\nu_2+1)$, we must prove that for large n,

$$\sup_{f\in\mathscr{C}_1,\, g\in\mathscr{C}_2} E_{f,g}\{\hat{g}^{(t)}(x_0) - g^{(t)}(x_0)\}^2 \ge Cn^{-4(\nu_2-t)/(2\nu_2+1)}.$$

This inequality follows from

$$\sup_{f\equiv 0,\, g\in\mathscr{C}_2} E_{f,g}\{\hat{g}^{(t)}(x_0) - g^{(t)}(x_0)\}^2 \ge Cn^{-4(\nu_2-t)/(2\nu_2+1)}, \tag{3.6}$$

which is true for all $\nu_2 > t$. To prove (3.6), note that when $f \equiv 0$ our model entails $Y_{ij}^2 = g(j/n) + \eta_{ij}$, where $\eta_{ij} \equiv g(j/n)(\varepsilon_{ij}^2 - 1)$. This is a replicated regression model, having mean function g and residuals with uniformly bounded variance. Techniques of Stone (1980), giving lower bounds to convergence rates for nonreplicated regression models, are easily modified to produce (3.6).

When $\nu_1 < \nu_2/(\nu_2+1)$, we must show that for large n,

$$\sup_{f\in\mathscr{C}_1,\, g\in\mathscr{C}_2} E_{f,g}\{\hat{g}^{(t)}(x_0) - g^{(t)}(x_0)\}^2 \ge Cn^{-4\nu_1(\nu_2-t)/\{(\nu_1+1)\nu_2\}}. \tag{3.7}$$

Our first proof of this inequality is valid for

$$\nu_1 < \nu_2/(\nu_2+1), \qquad \nu_2 > \max(t,1),\ t = 0,1,\dots. \tag{3.8}$$

The only case of interest not covered by these conditions is

$$\nu_1 < \nu_2/(\nu_2+1), \qquad 0 < \nu_2 \le 1,\ t = 0, \tag{3.9}$$

and we shall treat this separately at the end of our main proof.

Assume condition (3.8). Let ψ_1, ψ_2 be real-valued functions having at least $\langle\nu_2\rangle + 2$ bounded derivatives on $(-\infty,\infty)$, such that ψ_1 vanishes outside $[0,1]$, ψ_2 vanishes outside $[-1,1]$, $\psi_1(\frac{1}{2}) \ne 0$, $\psi_2^{(t)}(0) \ne 0$ and $\sup|\psi_j^{(i)}| \le \frac{1}{2}B$ for $0 \le i \le \langle\nu_2\rangle + 2$ and $j = 1,2$. Fix $c > 0$ and put $m_1 \equiv [cn^{\nu_1/(\nu_1+1)}]$, $m \equiv [n^{1-(2\nu_1)/\{(\nu_1+1)\nu_2\}}/m_1]$ (where $[x]$ denotes the integer part of x), $m_2 \equiv m_1 m$ and $\delta_i \equiv m_i/n$ for $i = 1,2$. Let m_0 be an integer such that $m_0 m_1 \le n$ and $m_0 \sim n/m_1$. Since we are assuming $\nu_2 > 1$, then $\nu_2/(\nu_2+1) < \frac{1}{2}\nu_2$, and so the hypothesis $\nu_1 < \nu_2/(\nu_2+1)$ entails $\nu_1 < \frac{1}{2}\nu_2$. This implies $m \to \infty$ as $n \to \infty$. Let $\{I_{ij},\ -m_0 \le i \le m_0 - 1 \text{ and } -m \le j \le m-1\}$ be a sequence of ± 1's, put $A(x,y) \equiv \delta_1^{\nu_1}\psi_1(x/\delta_1)\psi_1(y/\delta_2)$, and define $f = f(\cdot\,|\{I_{ij}\})$ by

$$f(x,y) = I_{ij}A(x - n^{-1}m_1 i,\ y - n^{-1}m_1 j) \quad \text{if } (x,y) \in \mathscr{I}_{ij},$$

$$f(x,y) = 0 \quad \text{if } (x,y) \notin \bigcup_{ij}\mathscr{I}_{ij},$$

where $\mathscr{I}_{ij} \equiv [n^{-1}m_1 i,\ n^{-1}m_1(i+1)) \times [n^{-1}m_1 j,\ n^{-1}m_1 j,\ n^{-1}m_1(j+1))$ for $-m_0 \le i \le m_0 - 1$ and $-m \le j \le m-1$, and where $\bigcup_{ij}$ denotes the union over these values of i, j. Write $\mathscr{F}$ for the class of all such fs. Let $G(x) \equiv \delta_1^{2\nu_1}\psi_2(x/\delta_2)$, $g_0 \equiv 1$, $g_1 \equiv (1-G)^{-1}$ and $\mathscr{G} \equiv \{g_0, g_1\}$. For large n, $\mathscr{F} \subseteq \mathscr{C}_1$ and $\mathscr{G} \subseteq \mathscr{C}_2$, provided $B > 1$. (The latter restriction may be removed at the cost of notational complexity.)

Let $\hat{g}^{(t)}(0)$ be any nonparametric estimator of $g^{(t)}(0)$. It suffices to show that

$$\liminf_{n\to\infty} n^{4\nu_1(\nu_2-t)/\{(\nu_1+1)\nu_2\}} \sup_{f\in\mathscr{F},\, g\in\mathscr{G}} E_{f,g}\{\hat{g}^{(t)}(0) - g^{(t)}(0)\}^2 > 0.$$

This result will follow if we prove that

$$\liminf_{n\to\infty} n^{4\nu_1(\nu_2-t)/\{(\nu_1+1)\nu_2\}} \sup_{g\in\mathcal{G}} E_g^*\{\hat{g}^{(t)}(0) - g^{(t)}(0)\}^2 > 0, \tag{3.10}$$

where E_g^* denotes expectation under the model

$$Y_{ij} = f\big(i/n, j/n|\{I_{\alpha\beta}\}\big) + g(j/n)^{1/2}\varepsilon_{ij}, \qquad -n \le i, j \le n,$$

in which the $I_{\alpha\beta}$'s are independent symmetric ± 1 variables [independent of the ε_{ij}'s, which are i.i.d. $N(0,1)$].

If (3.10) fails, choose a sequence $\{n_k\}$ such that the left-hand side of (3.10) converges to zero as $n \to \infty$ through $\{n_k\}$. Since

$$|g_0^{(t)}(0) - g_1^{(t)}(0)| \sim \delta_1^{2\nu_1}\delta_2^{-t}|\psi_2^{(t)}(0)| \sim \text{const.}\, n^{-2\nu_1(\nu_2-t)/\{(\nu_1+1)\nu_2\}},$$

then the decision rule $\hat{D}$ given by $\hat{D} = 0$ if $|\hat{g}^{(t)}(0) - g_0^{(t)}(0)| \le |\hat{g}^{(t)}(0) - g_1^{(t)}(0)|$, $\hat{D} = 1$ otherwise, provides asymptotically perfect discrimination between $g_0^{(t)}(0)$ and $g_1^{(t)}(0)$ as $n \to \infty$ through $\{n_k\}$, in the sense that

$$P_{g_0}^*(\hat{D} = 1) + P_{g_1}^*(\hat{D} = 0) \to 0.$$

We shall complete our proof by showing that this is impossible, even for the likelihood ratio (LR) rule. It suffices to show that if the true g is g_0 then the chance that the LR rule picks g_1 is bounded away from zero as $n \to \infty$. We may confine attention to the LR rule based on $\{Y_{ij},\ |i| \le m_0 m_1$ and $|j| \le m m_1\}$. (Note that $m_0 m_1 \sim n$ and that $g_0(x) = g_1(x)$ for $|x| > m m_1/n$. Therefore, Y_{ij} with $|j| > m m_1$ provides no information for discriminating between g_0 and g_1.)

Let a, b, α, β be integers satisfying $-m_0 \le a \le m_0 - 1$, $-m \le \alpha \le m - 1$, $1 \le b$, $\beta \le m_1$. If $i = a m_1 + b$ and $j = \alpha m_1 + \beta$, write $Y_{ab\alpha\beta}$ and $\varepsilon_{ab\alpha\beta}$ for Y_{ij} and ε_{ij}, respectively. For fixed a, α, the likelihood of $\{Y_{ab\alpha\beta},\ 1 \le b, \beta \le m_1\}$ is proportional to

$$\left(\exp\left[-\frac{1}{2}\sum_b\sum_\beta \frac{\{Y_{ab\alpha\beta} + A(b/n, \beta/n)\}^2}{g_0\{(\alpha m_1 + \beta)/n\}}\right] + \exp\left[-\frac{1}{2}\sum_b\sum_\beta \frac{\{Y_{ab\alpha\beta} + A(b/n, \beta/n)\}^2}{g_1\{(\alpha m_1 + \beta)/n\}}\right]\right)\left[\prod_{\beta=1}^{m_1} g\left\{\frac{(\alpha m_1 + \beta)}{n}\right\}\right]^{-m_1/2}$$

The chance that the LR rule wrongly picks g_1, equals the probability that

$$\exp\left[-\frac{1}{2}\sum_a\sum_b\sum_\alpha\sum_\beta \frac{\varepsilon_{ab\alpha\beta}^2}{g_1\{(\alpha m_1 + \beta)/n\}}\right]\left\{\prod_{j=1}^{n} g_1\left(\frac{j}{n}\right)\right\}^{-n/2}$$

$$\times \prod_a\prod_\alpha\left(1 + \exp\left[-2\sum_b\sum_\beta \frac{\{A(b/n, \beta/n)^2 + A(b/n, \beta/n)\varepsilon'_{ab\alpha\beta}\}}{g_1\{(\alpha m_1 + \beta)/n\}}\right]\right).$$

$$\ge \exp\left(-\frac{1}{2}\sum_a\sum_b\sum_\alpha\sum_\beta \varepsilon_{ab\alpha\beta}^2\right)\prod_a\prod_\alpha\left(1 + \exp\left[-2\sum_b\sum_\beta\left\{A\left(\frac{b}{n}, \frac{\beta}{n}\right)^2 + A\left(\frac{b}{n}, \frac{\beta}{n}\right)\varepsilon_{ab\alpha\beta}\right\}\right]\right).$$

(Here we have used symmetry of the Normal distribution, which implies that $\varepsilon_{ab\alpha\beta}$ and $I_{\alpha\beta}\varepsilon_{ab\alpha\beta}$ have the same distribution.) Equivalently, since $G = 1 - g_1^{-1}$, it equals the chance that

$$\exp\left(n \sum_{j=1}^{n}\left[G\left(\frac{j}{n}\right) + \log\left\{1 - G\left(\frac{j}{n}\right)\right\}\right] + \sum_a \sum_b \sum_\alpha \sum_\beta \left(\varepsilon_{ab\alpha\beta}^2 - 1\right) G\left\{\frac{\alpha m_1 + \beta}{n}\right\}\right)$$
$$\times \prod_a \prod_\alpha \left\{\left(1 + \exp\left[-2\sum_b \sum_\beta \left\{A\left(\frac{b}{n}, \frac{\beta}{n}\right)^2 + A\left(\frac{b}{n}, \frac{\beta}{n}\right)\varepsilon_{ab\alpha\beta}\right\}\right.\right.\right.$$
$$\left.\left.\times\left\{1 - G\left(\frac{\alpha m_1 + \beta}{n}\right)\right\}\right]\right)^2$$
$$\left.\times\left(1 + \exp\left[-2\sum_b \sum_\beta \left\{A\left(\frac{b}{n}, \frac{\beta}{n}\right)^2 + A\left(\frac{b}{n}, \frac{\beta}{n}\right)\varepsilon_{ab\alpha\beta}\right\}\right]\right)^{-2}\right\} \geq 1.$$

Denote the left-hand side of this inequality by B and put

$$d_1 \equiv \sum_b \sum_\beta A\left(\frac{b}{n}, \frac{\beta}{n}\right)^2 \sim c^{2(\nu_1+1)}\left(\int \psi_1^2\right)^2 \equiv d,$$

$$N_{a\alpha} \equiv d_1^{-1/2} \sum_b \sum_\beta A\left(\frac{b}{n}, \frac{\beta}{n}\right)\varepsilon_{ab\alpha\beta} \overset{\mathscr{D}}{=} N(0,1).$$

Noting that $\delta_1 \sim m_0^{-1} \sim cn^{-1/(\nu_1+1)}$ and $\delta_2 \sim n^{-2\nu_1/\{(\nu_1+1)\nu_2\}}$, we see that

$$\log B = -\{1 + o(1)\}\frac{1}{2}n \sum_{j=1}^{n} G\left(\frac{j}{n}\right)^2 + O_p\left[\left\{n \sum_{j=1}^{n} G\left(\frac{j}{n}\right)^2\right\}^{1/2}\right]$$
$$+ \{1 + o_p(1)\}4\sum_a \sum_\alpha \exp\{-2(d_1 + d_1^{1/2}N_{a\alpha})\}$$
$$\times\left[1 + \exp\{-2(d_1 + d_1^{1/2}N_{a\alpha})\}\right]^{-1}(d_1 + d_1^{1/2}N_{a\alpha})G\left(\frac{\alpha m_1}{n}\right)$$
$$= -\{1 + o_p(1)\}\frac{1}{2}n^2\delta_1^{4\nu_1}\delta_2 \int \psi_2^2 + \{1 + o_p(1)\}8m_0\delta_1^{2\nu_1-1}\delta_2 s$$
$$= n^{\{\nu_2 - \nu_1(\nu_2+1)\}/\{(\nu_1+1)\nu_2\}}\left\{8c^{2\nu_1-2}s - \frac{1}{2}c^{4\nu_1}\int \psi_2^2 + o_p(1)\right\},$$

where $s \equiv (\int \psi_2)E([\exp\{2(d + d^{1/2}N)\} + 1]^{-1}(d + d^{1/2}N))$, $N \overset{\mathscr{D}}{=} N(0,1)$ and c is chosen so that the expectation is nonzero. Choose ψ_2 to be either nonnegative or nonpositive, the sign being selected so that $s > 0$, and choose $|\psi_2|$ so small that $8c^{2\nu_1-2}s - \frac{1}{2}c^{4\nu_1}\int\psi_2^2 > 0$. Then $B \to +\infty$ in probability, implying that the chance that the LR rule picks g_1 when g_0 is the true variance function converges to one as $n \to \infty$. This completes our proof in the presence of condition (3.8).

The proof when (3.9) holds is simpler. Adopt the same notation as before, except that m is redefined as m_0 ($\sim n/m_1$), $\psi_2 \equiv 1$, and m_2 and δ_2 are no longer needed. Pursue the same argument. □

We next sketch a proof of the fact that if (3.4) holds for a nonparametric estimator $\hat{z}_0$ of z_0, then $a_n \to \infty$. We treat only the case $\nu_1 < \nu_2/(\nu_2 + 1)$, which is the context of the major part of our proof of (3.3). The case $\nu_1 \geq \nu_2/(\nu_2 + 1)$ is similar. Our argument is almost identical to that employed to derive (3.3).

Assume that estimation takes place on $[-1, 1]^2$. Use the same class of f's but change g_0, g_1 from 1, $(1 - G)^{-1}$ respectively to H, $H + G$, respectively, where G is as before and H is a positive, strictly convex function with unique minimum interior to $[-1, 1]$. For definiteness we shall take $H(z) \equiv (1 + z^2)B_0$, where our selection of the positive constant B_0 depends on the value of B. Let z_{00} $(= 0)$ and z_{01} be the values which minimize g_0 and g_1, respectively. Now,

$$g_1'(z) = 2B_0 z + \delta_1^{2\nu_1}\delta_2^{-1}\psi_2'(z/\delta_2),$$

which equals zero when $z = z_{01}$. Thus, by appropriate choice of ψ_2 we may ensure that z_{00} and z_{01} are distant apart an amount which is asymptotic to const.$\delta_1^{2\nu_1}\delta_2^{-1}$. The argument given during our proof of (3.3) shows that it is impossible to discriminate between z_{00} and z_{01}, and so it is also impossible to discriminate between z_{00} and z_{01}. Therefore no nonparametric estimator of z_0 can converge to z_0 more rapidly than $\delta_1^{2\nu_1}\delta_2^{-1}$, and the latter is asymptotic to a constant multiple of

$$n^{-2\nu_1(\nu_2 - 1)/\{(\nu_1 + 1)\nu_2\}} = \max\left(n^{-2(\nu_2 - 1)/(2\nu_2 + 1)}, n^{-2\nu_1(\nu_2 - 1)/\{(\nu_1 + 1)\nu_2\}}\right),$$

the above identity holding since $\nu_1 < \nu_2/(\nu_2 + 1)$. It follows that if (3.4) holds, then $a_n \to \infty$.

A proof of the fact that (3.5) entails $a_n \to \infty$ is similar. It uses the same g_0 $(= H)$ and g_1 $(= H + G)$ as above, but has the class of f's changed from $\mathscr{F}$ to $\mathscr{F}' \equiv \{F + f\colon f \in \mathscr{F}\}$, where F is an appropriate bivariate function which is strictly monotone in both variables. For example, if $\tau_0 = 2$ and z_0 is close to zero, then we may take $F(x, z) \equiv (x + 1)^2 + (z + 1)^2$.

4. Random design case. Although the fixed design case is the more important, analogues of our results may be obtained if (x_i, z_i), $1 \leq i \leq N$, are random variables distributed within the square $[0, 1]^2$ according to density d, rather than points on a lattice. In the present section we briefly discuss the random design case. The reader is referred to Prakasa Rao (1983), Section 4.2 for details of nonparametric regression estimation.

Assume that N observations (x_i, Y_i, z_i) are generated by the model

$$Y_i = f(x_i, z_i) + g(z_i)^{1/2}\varepsilon_i, \qquad 1 \leq i \leq N,$$

where f is ν_1-smooth, g is ν_2-smooth, the density d of the pairs (x_i, z_i) is

$\max(\nu_1, \nu_2)$-smooth and, conditional on the (x_i, z_i)'s, the ε_i's are independent with zero mean and uniformly bounded second moments. A kernel estimator of d is

$$\hat{d}(x, z) \equiv \left(Nh_1^2\right)^{-1} \sum_{j=1}^{N} K_1\{(x_j - x)/h_1, (z_j - z)/h_1\}, \tag{4.1}$$

where K_1 is a compactly supported bivariate function as in Theorem 3.1 of Stute (1984) and such that $\int x^i z^j K_1(x, z)\, dx\, dz = 1$ if $i = j = 0$ and $= 0$ if $1 \le i + j \le \langle \nu_1 \rangle$. A kernel estimator of f is

$$\begin{aligned} \hat{f}(x, z) &\equiv \hat{s}(x, z)/\hat{d}(x, z), \\ \hat{s}(x, z) &= \left(Nh_1^2\right)^{-1} \sum_{j=1}^{N} Y_j K_1\{(x_j - x)/h_1, (z_j - z)/h_1\}. \end{aligned} \tag{4.2}$$

Let $\hat{d}_i(x, z)$ and $\hat{s}_i(x, z)$ be as in (4.1) and (4.2), but with the sums taken only over $j \neq i$, and let $\hat{f}_i(x, z) \equiv \hat{s}_i(x, z)/\hat{d}_i(x, z)$, $r_i \equiv Y_i - f(x_i, z_i)$ (not observable) and $\hat{r}_i \equiv Y_i - \hat{f}_i(x_i, z_i)$. Fix $0 < \delta < \frac{1}{2}$. Analogues of $\tilde{g}$ and $\hat{g}$ are

$$\begin{aligned} \tilde{g}(z) &\equiv \sum_{j=1}^{N} r_j^2 I(\delta < x_j < 1 - \delta) K_2\{(z_j - z)/h_2\} \\ &\qquad \Big/ \sum_{j=1}^{N} I(\delta < x_j < 1 - \delta) K_2\{(z_j - z)/h_2\}, \\ \hat{g}(z) &\equiv \sum_{j=1}^{N} \hat{r}_j^2 I(\delta < x_j < 1 - \delta) K_2\{(z_j - z)/h_2\} \\ &\qquad \Big/ \sum_{j=1}^{N} I(\delta < x_j < 1 - \delta) K_2\{(z_j - z)/h_2\}, \end{aligned} \tag{4.3}$$

respectively, where K_2 is a univariate function satisfying $\int z^i K_2(z)\, dz = 1$ for $i = 0$ and $= 0$ for $1 \le i \le \langle \nu_2 \rangle$.

Take $h_1 = N^{-1/(2\nu_1 + 2)}$ and write $a_N = N^{-\nu_1/(\nu_1 + 1)}$. By moment calculations applied to $\hat{s}^{(r,s)}$ and $\hat{d}^{(r,s)}$ for $0 \le r + s \le \nu_1$, using (4.2) we find that for $\delta \le x$, $z \le 1 - \delta$,

$$\left\{\hat{f}^{(r,s)}(x, z) - f^{(r,s)}(x, z)\right\}^2 = O_p\{N^{-(\nu_1 - r - s)/(\nu_1 + 1)}\}. \tag{4.4}$$

By Theorem 3.1 of Stute (1984),

$$\sup\left\{|\hat{d}_i(x, z) - d(x, z)| : 1 \le i \le N,\ \delta \le x,\ z \le 1 - \delta\right\} = O_p\left\{(a_N \log N)^{1/2}\right\}.$$

Assuming that d is bounded away from zero on $[0,1]^2$, one can show that, uniformly in $\delta < x$, $z < 1 - \delta$,

$$\begin{aligned} \hat{f}_i - f &= (\hat{s}_i - f\hat{d}_i)/d - (\hat{s}_i - f\hat{d}_i)(\hat{d}_i - d)/d^2 \\ &\quad + (\hat{s}_i - f\hat{d}_i)(\hat{d}_i - d)^2/(d^2 \hat{d}_i) + O_p(a_N \log N). \end{aligned} \tag{4.5}$$

Let $h_2 \to 0$ such that $Nh_2 \to \infty$. By moment calculations applied to each term in $\tilde{g}^{(t)}$, it follows that, for $0 \le t \le \langle \nu_2 \rangle$, $\delta \le z \le 1 - \delta$,

$$\{\tilde{g}^{(t)}(z) - g^{(t)}(z)\}^2 = O_p\Big\{\big(Nh_2^{2t+1}\big)^{-1} + h_2^{2(\nu_2 - t)}\Big\}. \tag{4.6}$$

Using (4.5), detailed calculations yield

$$\{\hat{g}^{(t)}(z) - g^{(t)}(z)\}^2 = O_p\Big\{\big(Nh_2^{2t+1}\big)^{-1} + h_2^{2(\nu_2 - t)} + h_2^{-2t}N^{-2\nu_1/(\nu_1+1)}\Big\}. \tag{4.7}$$

Equations (4.4), (4.6) and (4.7) are analogues of (2.6), (2.8) and (2.11), respectively.

We may also derive analogues of (1.3), (1.4) and (1.5), by following essentially the arguments given in Section 1. It is necessary to show that

$$\sup_{\delta < z < 1-\delta} |\hat{g}^{(2)}(z) - g^{(2)}(z)| \to 0, \qquad \sup_{\delta < x,\, z < 1-\delta} |\hat{f}^{(i,j)}(x, z) - f^{(i,j)}(x, z)| \to 0$$

in probability, where $(i, j) = (0, 1)$ or $(1, 0)$. The trick is to decompose $\hat{g}^{(2)}$ and $\hat{f}^{(i,j)}$ into a series of terms, each of which is a ratio of two consistent function estimators. Assuming sufficiently many moments of the errors ε_i and Hölder continuity of derivatives of K_1 and K_2, uniform consistency of these function estimators may be proved by using the "continuity argument"; see, for example, Stone [(1984), foot of page 1292] and Hall (1985). The technique is intricate and laborious, but conceptually straightforward. It gives the same rates of convergence exhibited in (1.3), (1.4) and (1.5), under the same conditions on f and g. Arguments similar to those in Section 3 may be employed to show that these rates are optimal.

Acknowledgments. The work of R. J. Carroll occurred during a visit to the Australian National University and was supported by the Air Force Office of Scientific Research. The paper has benefited from suggestions by a referee and an Associate Editor.

REFERENCES

BOX, G. E. P. (1988). Signal to noise ratios, performance criteria and transformation (with discussion). *Technometrics* **30** 1–40.

BUCKLEY, M. J., EAGLESON, G. K. and SILVERMAN, B. W. (1988). The estimation of residual variance in nonparametric regression. *Biometrika* **75** 189–199.

EDDY, W. F. (1980). Optimum kernel estimators of the mode. *Ann. Statist.* **8** 870–882.

GASSER, T., SROKA, L. and JENNER, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73** 625–633.

HALL, P. (1985). Asymptotic theory of minimum integrated square error for multivariate density estimation. In *Multivariate Analysis-VI* (P. R. Krishnaiah, ed.) 289–309. Academic, New York.

HALL, P. and CARROLL, R. J. (1989). Variance function estimation in regression: The effect of estimating the mean. *J. Roy. Statist. Soc. Ser. B* **51** 3–14.

LEON, R. V., SHOEMAKER, A. C. and KACKAR, R. N. (1988). Performance measures independent of adjustment: An explanation and extension of Taguchi's signal to noise ratios (with discussion). *Technometrics* **30** 253–285.

MÜLLER, H.-G. (1984). Smooth optimum kernel estimators of densities, regression curves and modes. *Ann. Statist.* **12** 766–774.

PARZEN, E. (1962). On estimation of a probability density and mode. *Ann. Math. Statist.* **35** 1065–1076.

PRAKASA RAO, B. L. S. (1983). *Nonparametric Functional Estimation.* Academic, New York.

STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

STONE, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *Ann. Statist.* **12** 1285–1297.

STUTE, W. (1984). The oscillation behavior of empirical processes: The multivariate case. *Ann. Probab.* **12** 361–379.

TAGUCHI, G. and WU, Y. (1985). *Introduction to Off-Line Quality Control.* Central Japan Quality Control Association, Tokyo.

DEPARTMENT OF STATISTICS
TEXAS A & M UNIVERSITY
COLLEGE STATION, TEXAS 77853

DEPARTMENT OF STATISTICS
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA, ACT 2601
AUSTRALIA